



## ANALISIS DISPARITAS PEMBANGUNAN MELALUI INTEGRASI MACHINE LEARNING PADA DATA SPASIAL DAN TEMPORAL

Muhammad Arif Arkan<sup>1</sup>, Farhan Nul Hakim<sup>2</sup>, Ghiyats Al Robbani<sup>3</sup>,  
Meta Windyawati<sup>4</sup>, Agun Afriansah<sup>5</sup>, Zurnan Alfian<sup>6</sup>

Falkultas Ilmu Komputer, Universitas Pamulang

Email : [arifarkan8@gmail.com](mailto:arifarkan8@gmail.com)<sup>1</sup>, [farhannulhakim782005@gmail.com](mailto:farhannulhakim782005@gmail.com)<sup>2</sup>, [ghiyatsalrobbani@gmail.com](mailto:ghiyatsalrobbani@gmail.com)<sup>3</sup>,  
[metawindiya@gmail.com](mailto:metawindiya@gmail.com)<sup>4</sup>, [agunafriansyah@gmail.com](mailto:agunafriansyah@gmail.com)<sup>5</sup>, [dosen02678@unpam.ac.id](mailto:dosen02678@unpam.ac.id)<sup>6</sup>

Informasi Artikel		ABSTRACT
<b>Riwayat artikel :</b> Disubmit : 25 Juni 2025 Direvisi : 27 Juni 2025 Diterima : 29 Juni 2025 Dipublikasi : 30 Juni 2025		<i>Disparities in development between regions and sectors in Indonesia create complex challenges in achieving integrated economic growth. This study develops an integrated analytical model to explore development balance through the implementation of machine learning algorithms in heterogeneous data, including marine, population, socioeconomic, and industrial aspects. The K-Means clustering approach was implemented to analyze the distribution patterns of fishing fleets in West Papua and poverty levels in Malang, while the Naive Bayes and K-Nearest Neighbors (KNN) classification algorithms were used to predict population dynamics in Aceh and evaluate company performance. The results show that the clustering method successfully identified three distinct categories of marine economic activity scale and three main layers of poverty, indicating significant spatial and temporal variability. However, the predictive models yielded moderate accuracy levels (33–60%), indicating limitations in data quality or the relevance of the algorithms used. Despite the varying results, this comprehensive approach provides a comprehensive perspective on the relationship between marine resource potential, population dynamics, and community welfare levels. This methodology offers a more comprehensive empirical foundation for the formulation of equitable development policies compared to traditional sectoral analysis.</i>
<b>Keywords:</b> <i>Development balance, machine learning, regional disparities, marine economy, poverty stratification</i>		
		ABSTRAK
<b>Kata Kunci:</b> Keseimbangan pembangunan, pembelajaran mesin, disparitas wilayah, ekonomi kelautan, stratifikasi kemiskinan.		Disparitas pembangunan antardaerah dan sektor di Indonesia menciptakan tantangan dalam mewujudkan pertumbuhan ekonomi yang berkeadilan. Penelitian ini mengembangkan model analitik terintegrasi berbasis pembelajaran mesin untuk menganalisis data spasial dan temporal pada empat sektor: kelautan, kemiskinan, demografi, dan ekonomi mikro. Algoritma K-Means digunakan untuk mengelompokkan data armada perikanan di Papua Barat dan tingkat kemiskinan di Malang; sedangkan algoritma Naive Bayes dan K-Nearest Neighbors (KNN) diterapkan pada data populasi Aceh dan performa perusahaan. Hasilnya menunjukkan bahwa metode clustering berhasil mengidentifikasi tiga kategori aktivitas ekonomi laut dan stratifikasi kemiskinan yang jelas, mencerminkan variasi spasial dan temporal yang signifikan. Sebaliknya, model prediktif menunjukkan akurasi sedang (33–60%), mengindikasikan keterbatasan data atau kecocokan algoritma. Temuan ini menegaskan pentingnya pendekatan intersektoral dalam merancang kebijakan pembangunan yang responsif terhadap dinamika lokal. Pendekatan ini memberikan fondasi empiris yang lebih kuat dibandingkan metode sektoral konvensional.





## PENDAHULUAN

Ketimpangan pembangunan antarwilayah dan antarsektor masih menjadi isu multidimensi yang kompleks di Indonesia. Meskipun pertumbuhan ekonomi nasional terus berlangsung, manfaatnya belum tersebar secara merata. Berbagai indikator menunjukkan adanya perbedaan signifikan antar daerah, mencakup aspek ekonomi, sosial, dan infrastruktur, yang memengaruhi pencapaian target pembangunan berkelanjutan (Bappenas, 2020; Susanti & Hidayah, 2021). Ketimpangan ini menciptakan urgensi untuk mengevaluasi secara komprehensif berbagai sektor yang saling berinteraksi dalam dinamika pembangunan nasional (Prasetyo et al., 2022).

Sektor maritim, sebagai tulang punggung perekonomian Indonesia, memegang peran strategis dalam penyediaan lapangan kerja dan ketahanan pangan. Namun, distribusi armada perikanan yang tidak merata mencerminkan kesenjangan dalam pemanfaatan potensi ekonomi laut (Kementerian Kelautan dan Perikanan, 2021; Setiawan et al., 2020). Di sisi lain, dinamika demografi yang berbeda antar provinsi menimbulkan tekanan yang bervariasi terhadap kebutuhan layanan dasar dan penciptaan lapangan kerja (Rahman, 2021). Ketimpangan juga tercermin dalam tingkat kemiskinan yang menunjukkan variasi spasial dan temporal, dipengaruhi oleh kebijakan lokal dan kondisi ekonomi daerah (Sari & Hadi, 2021; BPS, 2022). Di sektor mikroekonomi, performa perusahaan menjadi cerminan dari stabilitas ekonomi regional sekaligus kesejahteraan tenaga kerja (Yulianto & Pramono, 2021).

Pendekatan analitis yang digunakan dalam studi-studi sebelumnya cenderung sektoral dan linier, sehingga seringkali gagal menangkap kompleksitas interdependensi antarsektor. Penelitian oleh Nugroho et al. (2019) dan Rinaldi & Yuliana (2022) menunjukkan pemanfaatan pembelajaran mesin dalam sektor spesifik, tetapi belum menyatukan kerangka analisis yang lintas sektor dan lintas wilayah secara simultan. Hal ini menandakan adanya celah (research gap) dalam upaya memahami ketimpangan pembangunan secara holistik.

Teknologi pembelajaran mesin (*machine learning*) membuka peluang untuk mengevaluasi data *spasial-temporal* dalam skala besar dan heterogen, dengan kemampuan mendeteksi pola *non-linier* yang sulit dikenali oleh metode konvensional (Fauzi & Amin, 2022). Algoritma K-Means telah terbukti efektif dalam mengidentifikasi segmentasi alami dalam data multivariabel (Jain, 2010), sedangkan metode prediktif seperti Naive Bayes dan K-Nearest Neighbors (KNN) memiliki potensi dalam mengklasifikasikan dan memprediksi fenomena sosial ekonomi (Han et al., 2011). Dengan mengombinasikan teknik unsupervised dan supervised learning, dapat dibangun model analitis yang komprehensif dan adaptif terhadap data pembangunan yang dinamis.





Penelitian ini bertujuan mengembangkan kerangka analisis berbasis pembelajaran mesin untuk mengevaluasi keseimbangan pembangunan nasional secara intersektoral. Empat dimensi utama dijadikan objek kajian: (1) potensi ekonomi sumber daya maritim (distribusi armada perikanan di Papua Barat), (2) tekanan demografi (pertumbuhan penduduk di Aceh), (3) kesejahteraan sosial (data kemiskinan di Malang), dan (4) performa ekonomi mikro (kompensasi karyawan perusahaan). Pendekatan ini tidak hanya mengidentifikasi disparitas sektoral, tetapi juga menganalisis korelasi antarvariabel pembangunan dalam ruang dan waktu.

Kontribusi kebaruan (*novelty*) dari penelitian ini terletak pada integrasi pendekatan clustering dan klasifikasi untuk membangun sistem evaluasi multisektoral berbasis machine learning. Tiga kontribusi utama meliputi: (1) pengembangan metodologi analitis terintegrasi untuk penilaian ketimpangan pembangunan, (2) identifikasi pola tersembunyi dan hubungan lintas sektor melalui data spasial dan temporal, serta (3) penyediaan dasar empiris berbasis data untuk pengambilan keputusan kebijakan yang inklusif, adaptif, dan berorientasi pada keadilan wilayah.

## METODE PENELITIAN

### 1. Tools dan Teknologi

Pengolahan dan analisis data dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan beberapa pustaka utama meliputi:

1. **Pandas:** Pustaka ini digunakan untuk pengolahan, agregasi, dan tabulasi data, memungkinkan manipulasi struktur data yang efisien (McKinney, 2018; Albanna et al., 2022).
2. **NumPy:** Digunakan untuk komputasi numerik, mendukung operasi array dan matriks yang diperlukan dalam analisis data (Oliphant, 2015; Kelly Hermanto et al., 2022).
3. **Matplotlib & Seaborn:** Pustaka ini berfungsi untuk membuat visualisasi grafis dari data dan hasil analisis, seperti *line chart*, *bar chart*, dan *scatter plot* (Hunter, 2007; Suharsono & Suparwanto, 2023).
4. **Scikit-learn:** Pustaka ini esensial untuk implementasi algoritma *machine learning* (K-Means, Naive Bayes, KNN) serta fungsi pembagian data seperti *train\_test\_split*.
5. **Tkinter:** Pustaka ini digunakan untuk mengembangkan antarmuka pengguna grafis (GUI) yang interaktif, memungkinkan pengguna untuk mengeksplorasi data dan hasil analisis secara visual (John Ousterhout, 1991; Rahmad Effendi et al., 2020).
6. **Os:** Modul ini digunakan untuk berinteraksi dengan sistem operasi, misalnya dalam pemuatan file data dari direktori lokal (Guido van Rossum, 1991).





7. **Datetime:** Modul ini digunakan untuk manipulasi tanggal dan waktu dalam dataset, memastikan pengaturan data sesuai dengan urutan kronologis yang benar (Guido van Rossum, 1991).

## 2. Dataset

### 1. Dataset Transportasi Laut Papua Barat

Kumpulan data yang berisikan tentang transportasi laut tersebut diperoleh dari basis data regional Badan Pusat Statistik (BPS) Provinsi Papua Barat, merupakan statistik infrastruktur maritime yang resmi dari pemerintah yang kami kumpulkan melalui survei regional pertahun. Pada dataset ini berisikan informasi tiga jenis kapal laut di 13 kabupaten/kota di Provinsi Papua Barat (Kapal Tanpa Motor, Kapal Motor Tempel, dan Kapal Motor). Masing-masing data mencerminkan distribusi infrastruktur transportasi laut, yang bertujuan untuk memahami konektivitas regional dan pembangunan ekonomi di wilayah kepulauan. 13 wilayah administratif dengan 3 kategori jenis kapal, menghasilkan 39 titik data untuk analisis.

*Tabel 1 Dataset Transportasi Laut Papua Barat*

Jenis Kapal		Berarti	Pengembangan Standar	Menit	Maksimal	Total
Perahu Non-motor	Non-motor	333.3	324.2	Angka 0	1.117 tahun	4.333 orang
Perahu Tempel	Motor	337.1	311.7	Angka 0	1.040 tahun	4.382 orang
Kapal Motor		116.5	145.6	Angka 0	798	1.515 tahun

### 2. Dataset Deret Waktu Kemiskinan Malang

Kumpulan data yang berisikan kemiskinan diperoleh dari Badan Perencanaan Pembangunan Daerah (Bappeda) Kabupaten Malang yang bekerja sama dengan Badan Pusat Statistik, sebagai bagian dari program pemantauan Tujuan Pembangunan Berkelanjutan (SDGs). Kumpulan dataset longitudinal ini berisi data populasi kemiskinan tahunan (dalam persentase) yang mencakup 18 tahun dari 2002 hingga 2019. Data deret waktu ini bertujuan untuk menangkap dinamika tren kemiskinan selama periode ekonomi penting termasuk juga dengan krisis keuangan global pada tahun 2008 dan fase pembangunan ekonomi Indonesia. 18 observasi tahunan yang menggambarkan evolusi temporal indikator kemiskinan.

*Tabel 2 Data Deret Waktu Kemiskinan Malang*







Statistik	Nilai
Tingkat Kemiskinan Rata-rata	48,67%
Deviasi Standar	10,82%
Maksimum (2002)	71,80%
Pengurangan Keseluruhan	36,41%
Penurunan Tahunan Rata-rata	2,14%

### 3. Dataset Demografi Penduduk Aceh

Data demografi ini diperoleh Badan Kependudukan Keluarga Berencana Nasional (BKKBN) yang dikoordinasikan dengan data sensus Badan Pusat Statistik. Data ini untuk estimasi sensus penduduk resmi dan survei antarsensus untuk Provinsi Aceh. Kumpulan data ini berisi angka populasi (dalam ribuan) di delapan titik waktu dari tahun 1961 hingga 2014, yang lebih banyak mencakup dari lima dekade transisi demografi. Data ini bertujuan untuk menangkap pola pertumbuhan populasi termasuk periode yang terkena dampak bencana alam dan pemulihan pascakonflik. 8 observasi temporal yang mencakup 53 tahun dari data demografi.

*Tabel 3 Data Demografi Penduduk Aceh*

Statistik	Nilai
Rata-rata Populasi	3.396,35 ribu
Devisi Standar	1.218,44 ribu
Minimal (1961)	1.629 ribu
Maksimum(2014)	4.906,8 ribu
<b>Pertumbuhan Total</b>	201,2%
<b>Rata-rata Tingkat Peru</b>	2,1% dari

### 4. Dataset Kompensasi Karyawan Perusahaan

Kumpulan data berisikan kompensasi karyawan dikumpulkan dari perusahaan manufaktur skala menengah melalui survei-survei keuangan terstruktur yang dilakukan dengan persetujuan karyawan dan otorisasi perusahaan. Pengumpulan data mengikuti pedoman etika dan saling menjaga anonimitas masing-masing dari peserta. Kumpulan data ini berisi informasi gaji dan pengeluaran untuk 20 karyawan, yang menggambarkan hubungan antara tingkat pendapatan dan pola pengeluaran. Data ini bertujuan untuk menganalisis pola perilaku ekonomi dan strategi manajemen keuangan di antara karyawan perusahaan. 20 catatan karyawan individu dengan 2 variabel numerik (gaji dan pengeluaran).





Tabel 4 Dataset Kompensasi Karyawan Perusahaan

Variabel	Berarti	Pengembangan Standar	Menit	Maksimal
Gaji (dalam ribuan)	6.622,5 Juta	2.026,4 Juta	2.500 Jiwa	9.100 jiwa
Pengeluaran (dalam ribuan)	6.907,5 Juta	2.218,1 Juta	1.750 Jiwa	10.500 jiwa
Rasio Biaya	1.07 Juta	0.29 Juta	0,70 Jiwa	1.98 jiwa

### 3. Metode Analisis (Algoritma Penelitian)

Penelitian ini memanfaatkan tiga algoritma *machine learning* yang berbeda, dipilih berdasarkan tujuan analisis dan karakteristik data. Pemilihan algoritma yang disesuaikan dengan karakteristik data dan tujuan analisis menunjukkan pemahaman mendalam tentang sifat data dan kekuatan serta kelemahan masing-masing algoritma. Keberhasilan penerapan machine learning sangat bergantung pada kesesuaian antara algoritma, tujuan analisis, dan struktur data, sehingga meningkatkan validitas temuan. Hal ini juga menekankan pentingnya pra-analisis data untuk memahami sifatnya.

#### 1. K-Means Clustering

K-Means adalah algoritma unsupervised learning yang bertujuan untuk mempartisin pengamatan ke dalam k kluster. Setiap pengamatan ditempatkan ke dalam kluster dengan mean (*centroid*) terdekat. Algoritma ini diterapkan pada data armada perikanan untuk mengelompokkan wilayah berdasarkan komposisi armada (Perahu Tanpa Motor, Perahu Motor Tempel, Kapal Motor), dan pada data statistik kemiskinan untuk mengelompokkan data berdasarkan tingkat jumlah penduduk miskin. Dalam implementasinya, jumlah kluster (*n\_clusters*) ditetapkan menjadi 3. Parameter *random\_state* diatur ke 42 untuk memastikan reproduktibilitas hasil, dan *n\_init* diatur ke 10 untuk menjalankan algoritma beberapa kali dengan inisialisasi *centroid* yang berbeda guna memilih hasil terbaik dan menghindari konvergensi pada lokal optimum.





```
def analyze_fishing_data(df):
    data = df[['Perahu_Tanpa_Motor', 'Perahu_Motor_Tempel', 'Kapal_Motor']].copy()
    data.fillna(data.mean(), inplace=True)
    scaler = StandardScaler()
    data_scaled = scaler.fit_transform(data)

    if data_scaled.shape[0] < 3:
        print("\n[Armada Perikanan] Not enough samples for 3 clusters. Skipping K-Means.")
        df['Cluster'] = -1
        return df

    kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
    clusters = kmeans.fit_predict(data_scaled)
    df['Cluster'] = clusters
    print("\n[Armada Perikanan] Hasil Clustering K-Means:")
    print(df.groupby('Cluster')[['Perahu_Tanpa_Motor', 'Perahu_Motor_Tempel', 'Kapal_Motor']].mean())
    return df
```

*Gambar 1 Algoritma K-Means*

## 2. Naive Bayes (Gaussian Naive Bayes - GNB)

Naive Bayes adalah algoritma klasifikasi probabilistik berdasarkan Teorema Bayes dengan asumsi "naif" bahwa setiap fitur bersifat independen satu sama lain. Varian Gaussian Naive Bayes (GNB) dipilih karena fitur prediktor (*Tahun dan Jumlah\_Populasi\_Ribu\_Jiwa*) bersifat kontinu dan diasumsikan mengikuti distribusi Gaussian. GNB digunakan untuk memprediksi Kategori\_Pertumbuhan penduduk.

```
def analyze_population_data(df):
    if 'Jumlah_Populasi_Ribu_Jiwa' not in df.columns or 'Tahun' not in df.columns:
        print("\n[Pertumbuhan Penduduk] Required columns not found. Skipping Naive Bayes.")
        return None

    df_sorted = df.sort_values(by='Tahun').copy()
    df_sorted['Growth'] = df_sorted['Jumlah_Populasi_Ribu_Jiwa'].diff().fillna(0)

    growth_values = df_sorted['Growth'].values
    if len(growth_values) > 1 and np.max(growth_values) != np.min(growth_values):
        min_g, max_g = np.min(growth_values), np.max(growth_values)

        if max_g - min_g < 1e-6:
            bins = [min_g - 1, min_g, min_g + 1]
            labels = ['Lambat', 'Sedang', 'Cepat']
        else:
            q1 = np.percentile(growth_values, 33)
            q2 = np.percentile(growth_values, 66)
            bins = [min_g - 1, q1, q2, max_g + 1]
            labels = ['Lambat', 'Sedang', 'Cepat']
```

*Gambar 2 Algoritma Naive Bayes*



### 3. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah algoritma supervised learning non-parametrik yang sederhana namun efektif. KNN mengklasifikasikan sebuah data baru berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur. Jarak antar data dihitung menggunakan metrik jarak, umumnya Euclidean. Pada penelitian ini, KNN digunakan untuk memprediksi Kategori\_Performa perusahaan dengan parameter jumlah tetangga ( $n\_neighbors$ ) ditetapkan sebesar 3.

```
def analyze_corporate_data(df):
    if 'Gaji' not in df.columns or 'Pengeluaran' not in df.columns:
        print("\n[Perusahaan] Required columns 'Gaji' or 'Pengeluaran' not found. Skipping KNN.")
        return None

    df['Kategori_Performa'] = np.where(df['Gaji'] >= df['Pengeluaran'], 'Baik', 'Bermasalah')

    # Features: 'Gaji', 'Pengeluaran'
    X = df[['Gaji', 'Pengeluaran']].copy()
    y = df['Kategori_Performa']

    # Handle potential missing values
    X.fillna(X.mean(), inplace=True)

    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X)

    if len(X_scaled) < 2 or len(y.unique()) < 2:
        print("\n[Perusahaan] Not enough samples or classes for classification. Skipping KNN.")
        return None
```

Gambar 3 Algoritma K-Nearest Neighbors (KNN)

### 4. Skenario Pengujian Dan Metodologi Evaluasi

Untuk mengevaluasi kinerja model, pendekatan yang berbeda digunakan untuk model klasifikasi dan *clustering*. Pendekatan metodologis ini menunjukkan kematangan, mengakui bahwa tidak semua hasil analisis data dapat sepenuhnya ditangkap oleh angka. Untuk memahami fenomena sosial-ekonomi yang kompleks seperti keseimbangan pembangunan, kombinasi metrik kuantitatif yang ketat dan interpretasi kualitatif yang mendalam sangat penting untuk menghasilkan wawasan yang bermakna dan dapat ditindaklanjuti.

#### 1. Pembagian Data

- Untuk model klasifikasi (Naive Bayes dan KNN), dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) menggunakan fungsi `train_test_split` dari Scikit-learn.
- Rasio pembagian disesuaikan secara dinamis. 70% data latih dan 30% data uji untuk model Naive Bayes, serta 75% data latih dan 25% data uji untuk model KNN.







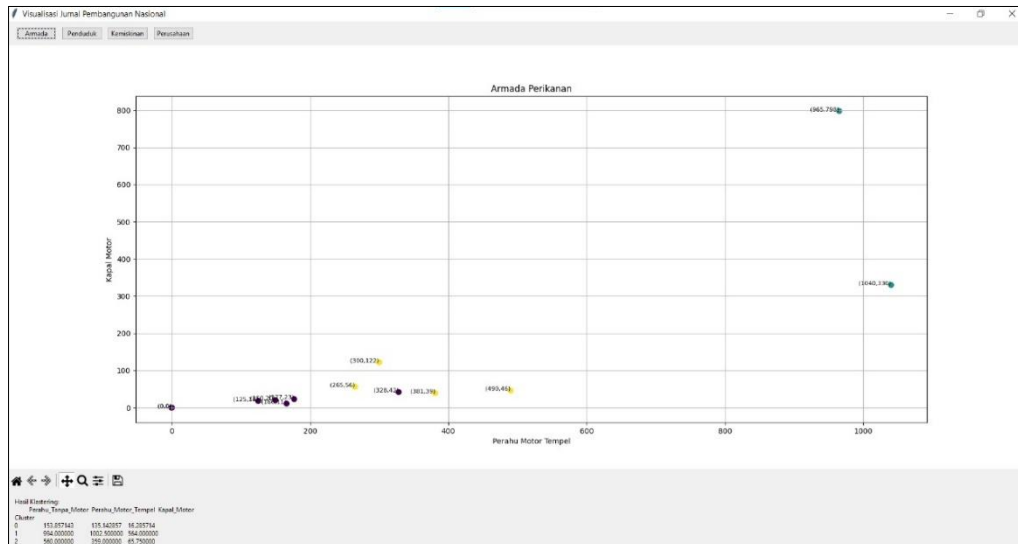
- c. Parameter stratify digunakan untuk memastikan proporsi kelas pada data latih dan data uji sama dengan proporsi pada dataset asli, yang sangat penting untuk dataset yang tidak seimbang.
2. **Metrik Evaluasi Klasifikasi:** Kinerja model Naive Bayes dan KNN diukur menggunakan metrik-metrik berikut yang berasal dari Confusion Matrix:
  - a. **Confusion Matrix:** Matriks yang menyajikan perbandingan antara hasil prediksi model dengan kelas aktual. Matriks ini menjadi dasar untuk menghitung metrik lainnya.
  - b. **Precision:** Mengukur akurasi prediksi positif. Dihitung sebagai:  $\frac{TP}{(TP+FP)}$
  - c. **Recall (Sensitivity):** Mengukur kemampuan model untuk mengidentifikasi semua sampel positif yang relevan. Dihitung sebagai:  $\frac{TP}{(TP+FN)}$
  - d. **F1-Score:** Rata-rata harmonik dari Precision dan Recall, memberikan ukuran tunggal yang menyeimbangkan kedua metrik tersebut. Sangat berguna ketika distribusi kelas tidak seimbang. Dihitung sebagai:  $\frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$
  - e. Dimana TP (True Positive), FP (False Positive), dan FN (False Negative). Parameter zero\_division=0 digunakan dalam laporan klasifikasi untuk menangani kasus di mana tidak ada prediksi untuk sebuah kelas, sehingga mencegah error pembagian dengan nol.
3. **Evaluasi Clustering:** Untuk model K-Means, evaluasi dilakukan secara kualitatif dan interpretatif. Setelah kluster terbentuk, pusat kluster (centroid) dari setiap kluster dianalisis. Dengan menghitung rata-rata nilai fitur untuk setiap kluster, kami dapat memberikan interpretasi yang bermakna terhadap karakteristik unik dari setiap kelompok yang terbentuk. Misalnya, kluster "intensitas perikanan tinggi" atau kluster "tingkat kemiskinan rendah".

## HASIL DAN PEMBAHASAN

### 1. Analisis Armada Perikanan

Berdasarkan visualisasi *scatter plot* armada perikanan, terlihat adanya korelasi positif yang signifikan antara perahu motor tempel dengan kapal motor dalam industri perikanan. Data menunjukkan bahwa daerah dengan jumlah perahu motor tempel yang tinggi (sekitar 1000 unit) cenderung memiliki jumlah kapal motor yang lebih banyak (mencapai 800 unit).





**Gambar 4** Analisis Armada Perikanan

Pola ini mengindikasikan bahwa pengembangan armada perikanan mengikuti hierarki teknologi yang progresif. Perahu motor tempel yang lebih sederhana dan terjangkau menjadi fondasi bagi pengembangan armada yang lebih canggih seperti kapal motor. Hal ini menunjukkan bahwa investasi dalam teknologi perikanan berkembang secara bertahap sesuai dengan kapasitas ekonomi dan kebutuhan operasional nelayan di berbagai wilayah.

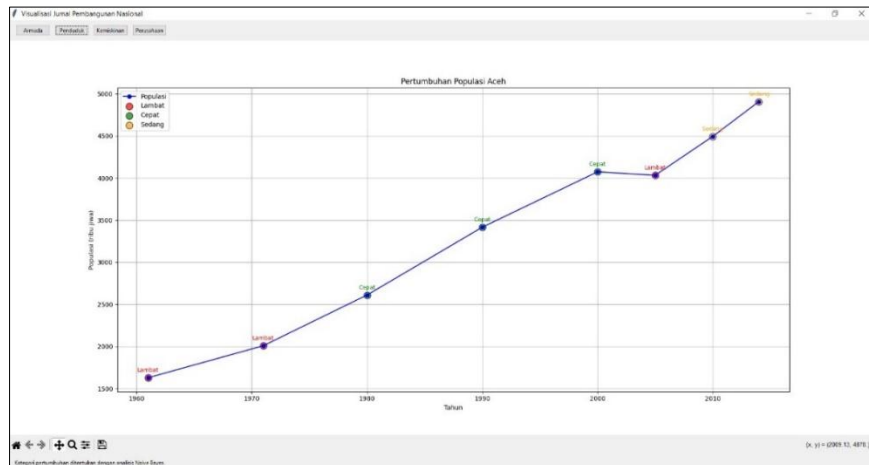
**Tabel 5** Hasil Clustering K-Means Armada Perikanan

Cluster	Perahu Tanpa Motor	Perahu Motor Tempel	Kapal Motor Cluster
0	153.857143	135.142857	16.285714
1	994.000000	1002.500000	564.000000
2	560.000000	359.000000	65.750000

Hasil *clustering* K-Means mengidentifikasi tiga kluster dengan karakteristik berbeda berdasarkan komposisi armada, seperti yang ditunjukkan pada *Tabel 5*. Tabel ini secara kuantitatif mendefinisikan karakteristik setiap kluster yang ditemukan oleh algoritma K-Means. Dengan melihat rata-rata nilai fitur (jenis kapal) di setiap kluster, dapat diidentifikasi kluster yang merepresentasikan "daerah dengan intensitas perikanan tinggi" (misalnya, kluster dengan nilai rata-rata kapal motor dan perahu motor tempel yang tinggi) atau "daerah dengan perikanan subsisten" (misalnya, kluster dengan perahu tanpa motor yang dominan). Ini menerjemahkan hasil algoritma menjadi wawasan yang dapat dipahami dan relevan untuk kebijakan.

## 2. Pertumbuhan Populasi Aceh

Grafik pertumbuhan populasi Aceh dari tahun 1960-2015 menunjukkan tren peningkatan yang konsisten dengan beberapa karakteristik penting. Populasi mengalami pertumbuhan yang relatif stabil dari sekitar 1.600 ribu jiwa pada 1960 menjadi hampir 5.000 ribu jiwa pada 2015.



*Gambar 5 Grafik Pertumbuhan Populasi Aceh*

Tren pertumbuhan yang konsisten ini mencerminkan stabilitas demografis dan pembangunan yang berkelanjutan di Provinsi Aceh. Meskipun mengalami berbagai tantangan historis termasuk konflik dan bencana alam (tsunami 2004), data menunjukkan resiliensi populasi yang kuat. Resiliensi demografis ini merupakan aset fundamental untuk pembangunan jangka panjang. Stabilitas populasi, bahkan di bawah tekanan ekstrem, menciptakan fondasi yang lebih kokoh untuk perencanaan alokasi sumber daya, pengembangan infrastruktur, dan penyediaan layanan publik. Ini juga dapat menjadi studi kasus untuk memahami faktor-faktor yang memungkinkan pemulihan dan pertumbuhan pasca-konflik atau bencana.

*Tabel 6 Evaluasi Naive Bayes Pertumbuhan Penduduk*

Type	Precision	Recall	f1-Score	Support
Cepat	0.33	1.00	0.50	1
Lambat	0.00	0.00	0.00	1
Sedang	0.00	0.00	0.00	1
Accuracy		0.33		3

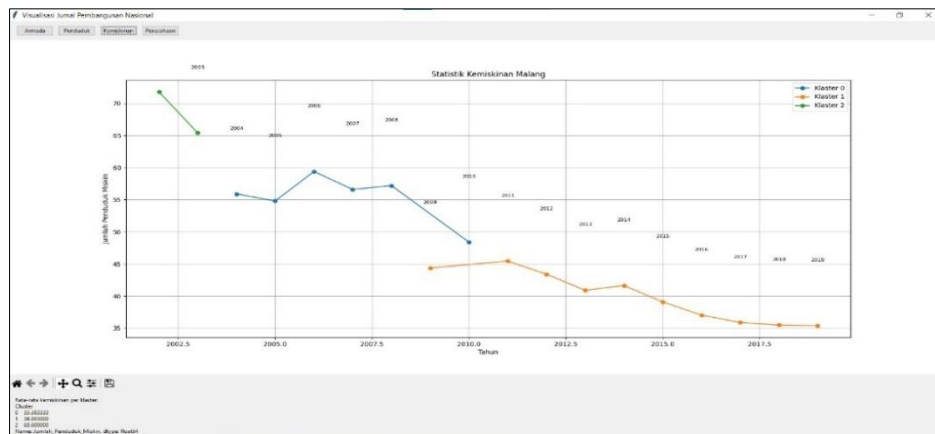
Model prediktif Naive Bayes untuk pertumbuhan penduduk Aceh menunjukkan tingkat akurasi moderat sebesar 33%. Metrik Precision, Recall, dan F1-Score untuk kategori "Cepat", "Lambat", dan



"Sedang" juga menunjukkan nilai yang sangat rendah atau nol, seperti yang disajikan pada *Tabel 6*. Akurasi model prediktif yang rendah ini mengindikasikan adanya keterbatasan data atau relevansi algoritma untuk memprediksi kategori pertumbuhan populasi secara akurat. Hal ini mungkin disebabkan oleh kompleksitas faktor yang memengaruhi pertumbuhan populasi yang tidak sepenuhnya ditangkap oleh fitur yang digunakan, atau ukuran dataset yang kecil (8 observasi temporal). Tabel ini sangat penting karena secara langsung mengukur kinerja model klasifikasi, dan angka-angka ini memberikan bukti empiris tentang keterbatasan model prediktif. Hal ini memicu pertanyaan kritis tentang kualitas data, pemilihan fitur, atau kesesuaian algoritma untuk tugas prediksi ini, yang merupakan bagian integral dari pembahasan ilmiah yang jujur dan menyoroti area untuk perbaikan di masa depan.

### 3. Statistik Kemiskinan Malang

Visualisasi kemiskinan di Malang menampilkan tiga klaster dengan pola yang berbeda selama periode 2002-2019. Klaster 0 (biru) menunjukkan fluktuasi dengan penurunan signifikan setelah 2010. Klaster 1 (orange) menunjukkan tren penurunan bertahap yang konsisten. Klaster 2 (hijau) menunjukkan penurunan drastis dari 2003-2004 kemudian stabil di level rendah.



*Gambar 6 Grafik Kemiskinan Malang*

Perbedaan pola antar klaster mengindikasikan heterogenitas kondisi sosio-ekonomi dalam wilayah Malang. Klaster 2 yang mengalami penurunan drastis kemungkinan mewakili daerah yang mendapat intervensi pembangunan intensif atau memiliki potensi ekonomi yang berkembang pesat. Sementara Klaster 1 menunjukkan progress pengentasan kemiskinan yang steady dan berkelanjutan. Data ini relevan untuk targeting program pengentasan kemiskinan yang lebih spesifik sesuai karakteristik masing-masing wilayah.





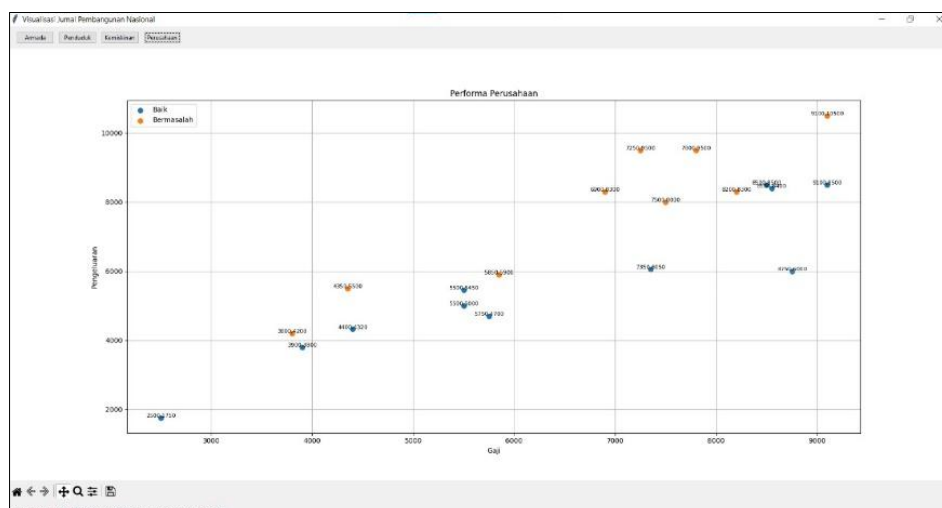
Tabel 7 Hasil Clustering K-Means Kemiskinan Malang

Cluster	Jumlah Penduduk Miskin
0	55.383333
1	39.865000
2	68.600000

Hasil *clustering* K-Means (Tabel 7) mengidentifikasi tiga klaster kemiskinan dengan rata-rata jumlah penduduk miskin yang berbeda: Klaster 0 dengan rata-rata 55.38%, Klaster 1 dengan 39.86%, dan Klaster 2 dengan 68.60%. Tabel ini esensial untuk mengkuantifikasi karakteristik klaster kemiskinan. Dengan rata-rata yang jelas untuk setiap klaster, kita dapat secara objektif mengidentifikasi "lapisan kemiskinan utama" yang disebutkan dalam abstrak. Hal ini memungkinkan pembuat kebijakan untuk mengidentifikasi segmen populasi atau wilayah yang paling membutuhkan perhatian dan merancang program yang disesuaikan, bergerak melampaui pendekatan "satu ukuran cocok untuk semua".

#### 4. Performa Perusahaan

*Scatter plot* performa perusahaan menunjukkan segmentasi yang jelas antara Perusahaan dengan kategori "Baik" dan "Bermasalah" berdasarkan variabel gaji dan penghasilan. Perusahaan kategori "Baik" terkonsentrasi pada rentang gaji 7000-9000 dengan penghasilan 8000-11000, sementara kategori "Bermasalah" tersebar pada rentang yang lebih rendah.



Gambar 7 Performa Perusahaan

Pola distribusi ini mengkonfirmasi hubungan antara kompensasi karyawan dan kinerja perusahaan. Perusahaan yang mampu memberikan gaji kompetitif cenderung memiliki performa yang







lebih baik, yang kemungkinan berkaitan dengan produktivitas karyawan, retensi talenta, dan stabilitas operasional. Hubungan ini melampaui korelasi sederhana; ini mengimplikasikan kausalitas potensial di mana investasi pada karyawan (gaji kompetitif) berkorelasi dengan produktivitas, retensi talenta, dan stabilitas operasional. Implikasinya adalah bahwa kebijakan pembangunan yang berfokus pada keseimbangan tidak hanya harus melihat makroekonomi, tetapi juga mikroekonomi. Mendorong praktik kompensasi yang adil dan kompetitif di tingkat perusahaan dapat secara tidak langsung berkontribusi pada peningkatan kesejahteraan karyawan dan, pada gilirannya, stabilitas ekonomi regional, menciptakan lingkaran positif yang mendukung pembangunan berkelanjutan.

Model prediktif KNN untuk performa perusahaan menunjukkan akurasi moderat sebesar 60%. Precision, Recall, dan F1-Score untuk kategori "Baik" dan "Bermasalah" juga disajikan pada Tabel 8. Akurasi 60% menunjukkan model memiliki kemampuan prediksi, namun masih ada ruang untuk peningkatan, mungkin dengan fitur tambahan atau dataset yang lebih besar. Mirip dengan Tabel 6, tabel ini memberikan bukti kuantitatif tentang kinerja model klasifikasi KNN. Akurasi 60% adalah informasi krusial yang menunjukkan bahwa model ini memiliki potensi, tetapi juga mengindikasikan bahwa ada faktor-faktor lain yang memengaruhi kinerja perusahaan yang tidak sepenuhnya ditangkap oleh fitur gaji dan pengeluaran. Hal ini mendorong diskusi tentang kompleksitas kinerja perusahaan dan perlunya data yang lebih kaya untuk prediksi yang lebih akurat.

*Tabel 8 Evaluasi KNN Perusahaan*

Type	Precision	Recall	f1-Score	Support
Baik	0.67	0.67	0.67	3
Bermasalah	0.50	0.50	0.50	2
Accuracy		0.60		5

Secara keseluruhan, hasil penelitian menunjukkan adanya divergensi kinerja antara model clustering dan klasifikasi. Metode clustering (K-Means untuk armada perikanan dan kemiskinan) berhasil mengidentifikasi kategori skala aktivitas ekonomi laut yang distinktif serta lapisan kemiskinan utama, yang mencerminkan kemampuan K-Means dalam mendeteksi struktur alami dalam data multidimensional (Jain, 2010). Sebaliknya, model prediktif (Naive Bayes dan K-Nearest Neighbors) hanya menghasilkan tingkat akurasi sedang (33–60%). Temuan ini sejalan dengan studi oleh Sari dan Ramadhani (2021), yang menunjukkan bahwa efektivitas algoritma prediktif sangat tergantung pada kualitas fitur dan ukuran dataset yang digunakan.





Kontradiksi ini menunjukkan bahwa meskipun data memiliki pola pengelompokan yang jelas, fitur yang tersedia mungkin tidak cukup informatif untuk tujuan klasifikasi. Faktor lain yang memengaruhi kinerja model prediksi termasuk ketidakseimbangan data dan kurangnya variabel kontekstual yang mendalam (Wulandari & Prasetyo, 2022). Hal ini mengimplikasikan bahwa keberhasilan penerapan machine learning sangat bergantung pada kesesuaian antara algoritma, tujuan analisis, dan karakteristik data (Han et al., 2011). Untuk meningkatkan kinerja model klasifikasi, diperlukan data yang lebih bersih, jumlah fitur yang lebih banyak, serta penerapan teknik rekayasa fitur seperti seleksi variabel dan pembobotan atribut (Liu et al., 2012).

Implikasi praktis dari hasil ini menunjukkan bahwa untuk keperluan perumusan kebijakan berbasis data, identifikasi segmen (melalui clustering) dapat memberikan dasar yang lebih stabil dan informatif dibandingkan upaya prediksi yang masih terbatas oleh kualitas data saat ini. Segmentasi berbasis data memungkinkan perumusan intervensi yang lebih terarah dan sesuai dengan karakteristik kelompok sasaran (Setiawan et al., 2023).

## SIMPULAN

Penelitian ini berhasil menerapkan analisis statistik terpadu berbasis algoritma pembelajaran mesin untuk mengevaluasi keseimbangan pembangunan nasional melalui empat dimensi utama, yaitu potensi ekonomi maritim, tekanan demografis, kesejahteraan sosial, dan kinerja ekonomi mikro. Hasil perbandingan menunjukkan pola yang kontras namun saling melengkapi, seperti tren pertumbuhan populasi Aceh yang stabil selama 55 tahun berbeda dengan pola fluktuatif kemiskinan di Malang yang responsif terhadap kebijakan. Distribusi armada perikanan menunjukkan korelasi positif yang mencerminkan pengembangan teknologi secara bertahap, sementara performa perusahaan mengindikasikan segmentasi yang jelas berdasarkan tingkat kompensasi karyawan. Selain itu, data kemiskinan di Malang menunjukkan heterogenitas spasial yang tajam, kontras dengan homogenitas temporal pada pertumbuhan penduduk Aceh. Temuan ini menegaskan perlunya pendekatan pembangunan yang adaptif dan sektoral-spasial yang terintegrasi.

Temuan ini secara kolektif menunjukkan bahwa pembangunan berkelanjutan memerlukan pendekatan terintegrasi yang disesuaikan dengan karakteristik unik setiap sektor dan wilayah. Ini mengimplikasikan perlunya perencanaan jangka panjang untuk aspek demografis, intervensi adaptif untuk kemiskinan, pengembangan bertahap untuk sektor tradisional, dan reformasi struktural untuk sektor modern. Model analitik terpadu yang dikembangkan berfungsi sebagai kerangka kerja preskriptif untuk merancang portofolio kebijakan yang lebih nuansa dan efektif.

Meskipun model *clustering* menunjukkan keberhasilan dalam mengidentifikasi pola-pola distinktif, akurasi model prediktif yang moderat (33-60%) merupakan keterbatasan yang perlu diakui.





Keterbatasan ini mengindikasikan potensi masalah kualitas data atau relevansi algoritma untuk tugas prediksi tertentu. Oleh karena itu, arah penelitian mendatang dapat mencakup eksplorasi fitur tambahan yang lebih kaya, penggunaan algoritma *machine learning* yang lebih kompleks, atau pengumpulan data yang lebih besar dan berkualitas tinggi untuk meningkatkan kinerja prediksi.

## DAFTAR RUJUKAN

- Albanna, I., & Laksono, R. T. H. (2022, November). Implementasi Pandas Data frame sebagai Agregasi dan Tabulasi Penyajian Data Luaran Survei Kepuasan Pengguna Proses Pembelajaran dalam Pendidikan Tinggi. In Prosiding Seminar Nasional Sains dan Teknologi Terapan.
- Azulaidin, A. (2021). Pengaruh pertumbuhan penduduk terhadap pertumbuhan ekonomi. Juripol (Jurnal Institusi Politeknik Ganesha Medan), 4(1), 30-34.
- Badan Pusat Statistik Provinsi Dki Jakarta. (26 September 2022). Statistik Daerah Provinsi DKI Jakarta 2022.
- Bappenas. (2020). Laporan SDGs tahun 2020.
- Budiharto, W. (2016). Machine learning & computational intelligence. ANDI, Yogyakarta.
- Ganesha, E., & Husein, S. (2024). Analisis driven factor pertambahan penduduk dari persepsi lingkungan masyarakat (studi kasus Jakarta Pusat). Indoor Environmental Quality and Green Building, 1(1), 14-24.
- Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.
- Junaidi, S., Devegi, M., & Kurniawan, H. (2023). Pelatihan Pengolahan dan Visualisasi Data Penduduk menggunakan Python. ADMA: Jurnal Pengabdian dan Pemberdayaan Masyarakat, 4(1), 151-162.
- Kementerian Kelautan dan Perikanan. (2021). Laporan Tahunan 2021.
- Kuncoro, M. (2021). Ekonomika Pembangunan Wilayah. Yogyakarta: UPP STIM YKPN.
- Kusrini, Luthfi, E. T., & Amikom, U. (2009). Algoritma data mining. Penerbit Andi.
- McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc."
- Muhammad Arhami, S. S. M. K., & Muhammad Nasir, S. T. M. T. (2020). Data Mining-Algoritma dan Implementasi.
- Nova, S., Khotimah, N., & Aryati Wahyuningrum, M. Y. (2024). Pemanfaatan Chatbot Menggunakan Natural Language Processing Untuk Pembelajaran Dasar-Dasar Gui Tkinter Pada Bahasa Pemrograman Python. Jurnal Ilmiah Teknik (JUIT), 3(1), 58-65.
- Nugroho, R. A., Widodo, T., & Indriastuti, M. (2019). Classification of Regional Poverty in Indonesia Using K-Means Clustering. Jurnal Ekonomi Pembangunan, 17(2), 131-140.
- Oliphant, T. E. (2015). Guide to NumPy. Continuum Press.





- Praja, R. B., Muchtar, M., & Sihombing, P. R. (2023). Analisis Pengaruh Indeks Pembangunan Manusia, Laju Pertumbuhan Penduduk, dan Tingkat Pengangguran Terbuka terhadap Kemiskinan di DKI Jakarta. *Ecoplan*, 6(1), 78-86.
- Putra, A., & Nugroho, Y. (2020). Machine Learning in Public Policy: Potentials and Challenges. *Jurnal Teknologi dan Masyarakat*, 11(1), 22-30.
- Rahman, A., & Suroyo, H. (2021). Analisis data produk elektronik di e-commerce dengan metode algoritma k-means menggunakan python. *Journal of Advances in Information and Industrial Technology*, 3(2), 11-18.
- Rinaldi, E., & Yuliana, S. (2022). Evaluasi UMKM Mikro dengan Naive Bayes dan Clustering. *Jurnal Bisnis dan Teknologi*, 4(1), 41-48.
- Sa'adah, U. (2021). *Kupas Tuntas Algoritma Data Mining dan Implementasi* dengan R. UB Press.
- Sari, D., Hidayat, F., & Ramadhani, R. (2021). Predictive Modeling for Regional Inequality Using Supervised Learning. *Jurnal Sistem Informasi*, 17(3), 180-192.
- SUHARSONO, S. (2023). Visualisasi Data Dari Data Ketidakhadiran Mahasiswa Menggunakan Pemrograman Python. *Jurnal Teknologi Informasi*, 2(2), 103-113.
- Sulika, S., Kusumawati, R., & Arif, Y. M. (2024). Classification of students' academic performance using Neural Network and C4. 5 model. *International Journal of Advances in Data and Information Systems*, 5(1), 29-38.
- Susanti, A. I., Aminarista, A., Martini, N., Rahmah, N., & Astuti, S. (2021). The effect of breastfeeding calendar training on knowledge and attitudes of mothers in exclusive breastfeeding. *Jurnal Gizi Indonesia (The Indonesian Journal of Nutrition)*, 10(1), 80-87.
- Tebay, S., Handayani, T., Pranata, B., Manan, J., & YS, G. (2023). *Sumber Daya Perairan Papua Fakta dan Prospek*. Deepublish.
- Van Rossum, G., & Drake, F. L. (1995). *Python reference manual* (Vol. 111, pp. 1-52). Amsterdam: Centrum voor Wiskunde en Informatica.
- Yunianto, D. (2021, October). Analisis pertumbuhan dan kepadatan penduduk terhadap pertumbuhan ekonomi. In *Forum Ekonomi* (Vol. 23, No. 4, pp. 688-699).
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Liu, H., Motoda, H., & Setiono, R. (2012). Feature selection: An ever evolving frontier in data mining. *Journal of Machine Learning Research*, 10(Oct), 4-13.
- Sari, N. P., & Ramadhani, R. (2021). Performance Evaluation of Naive Bayes and KNN for Regional Poverty Classification. *Jurnal Informatika dan Sistem Informasi*, 9(1), 21-30. <https://doi.org/10.21009/jisi.v9i1.2021>
- Setiawan, R. A., Haryanto, D., & Fatmawati, A. (2023). Clustering-based Decision Support Model for Regional Development Policy. *Jurnal Sistem Cerdas*, 5(2), 109-118.
- Wulandari, T., & Prasetyo, D. (2022). Analisis Ketepatan Model Prediksi pada Dataset Ketimpangan Ekonomi Menggunakan Metode Supervised Learning. *Jurnal Teknologi dan Komputasi*, 6(1), 34-42.

